COMPUTER VISION
CS-6350

Prof. Sukhendu Das
Deptt. of Computer Science and Engg.,
IIT Madras, Chennai – 600036.

Email: sdas@iitm.ac.in
URL: //www.cse.iitm.ac.in/~sdas
//www.cse.iitm.ac.in/~vplab/computer_vision.html

Jan – 2016.

INTRODUCTION

Contents to be covered

1 Introduction
2 Neighborhood and Connectivity of pixels
3 Fourier Theory, Filtering in spatial and spectral domains
4 3D transformations, projection and stereo
5 Histogram based image processing
6 Concepts in Edge Detection
7 Hough Transform
8 Scale-Space - Image Pyramids
9 Feature extraction (recent trends) – detectors and descriptors
10 Image segmentation
11 Texture analysis using Gabor filters
12 Pattern Recognition
13 Bag of Words and Prob. Graphical Models
14 Object Recognition
15 Motion Analysis
16 Shape from Shading
17 Wavelet transform
18 Reconstruction - affine, model-based
19 Registration and Matching

20 Solid Modelling; 21. Color
22. Hardware; 23. Morphology

Use slides as brief :
Points, comments, links

These are not substitute for materials in books

References

1. "Digital Image Processing"; R. C. Gonzalez and R. E. Woods; Addison Wesley; 1992+.

2. "3-D Computer Vision"; Y. Shirai; Springer-Verlag, 1984.

3. "Digital Image Processing and Computer Vision"; Robert J. Schallkoff; John Wiley and Sons; 1989+.

4. "Pattern Recognition: Statistical. Structural and Neural Approaches"; Robert J. Schallkoff; John Wiley and Sons; 1992+.

5. "Computer Vision: A Modern Approach"; D. A. Forsyth and J. Ponce; Pearson Education; 2003.

6. "Computer Vision: Algorithms and Applications" by Richard Szeliski; Springer-Verlag London Limited 2011 .

7. Multiple View geometry. R. Hartley and A. Zisserman. 2002 Cambridge university Press

References (Contd..)

Journals:

- IEEE-T-PAMI (Transactions on Pattern Analysis and Machine Intelligence)
- IEEE-T-IP (Transactions on Image processing)
- PR (Pattern Recognition)
- PRL (Pattern Recognition Letters)
- CVGIP (Computer Vision, Graphics & Image Processing)
- IJCV (International Journal of Computer Vision)

Online links

1. CV online: **http://homepages.inf.ed.ac.uk/rbf/CVonline**
2. Computer Vision Homepage:

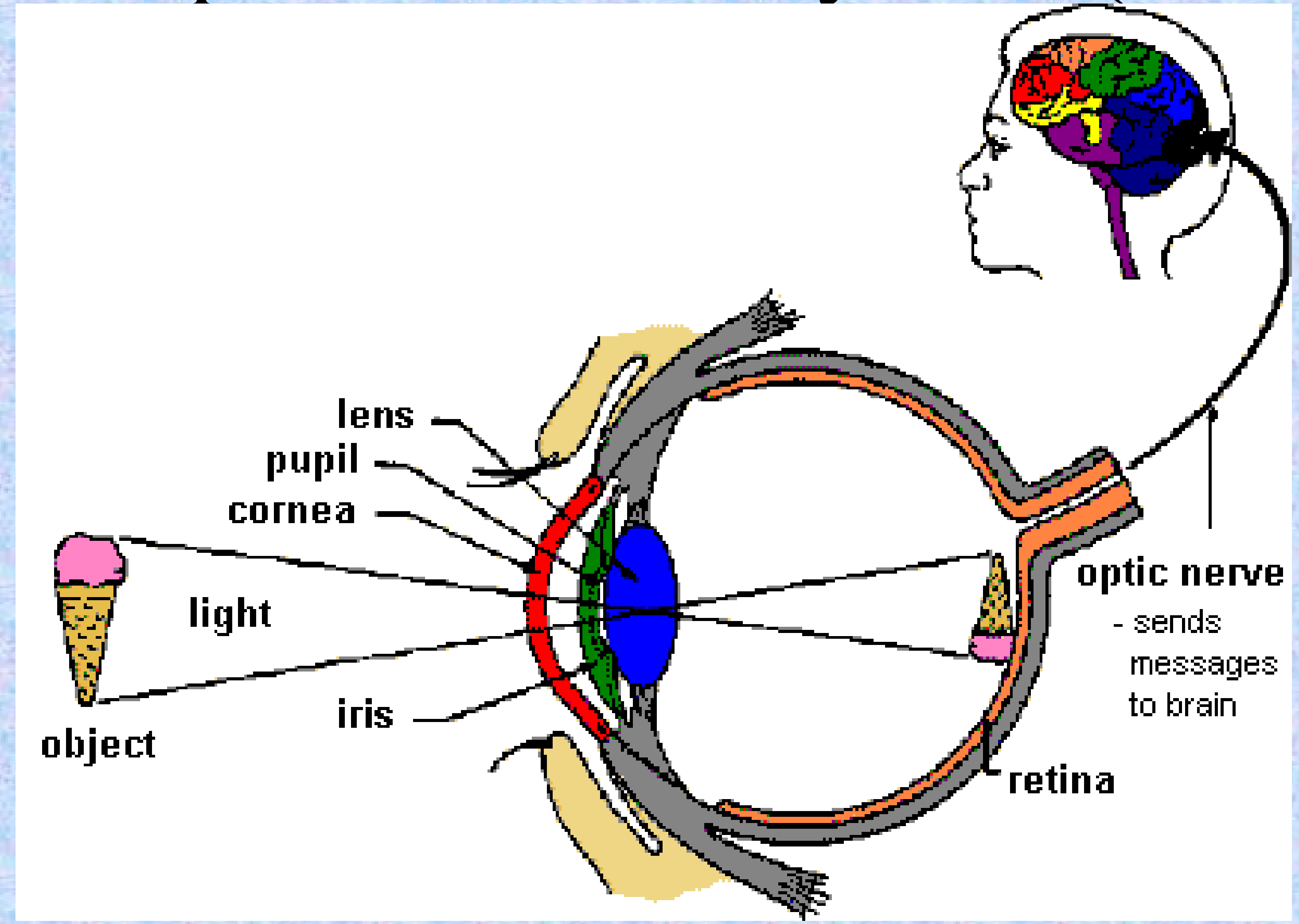
http://www-2.cs.cmu.edu/afs/cs/project/cil/ftp/html/vision.html

Typical Distribution of marks for Evaluation/grading

Quiz (50 mins.) -	**15 - 20**
End Sem exam (120-150 mins.) -	**35 – 40**
TPA -	**35 - 40**
TUTS -	**05 - 10**
Total	**100**

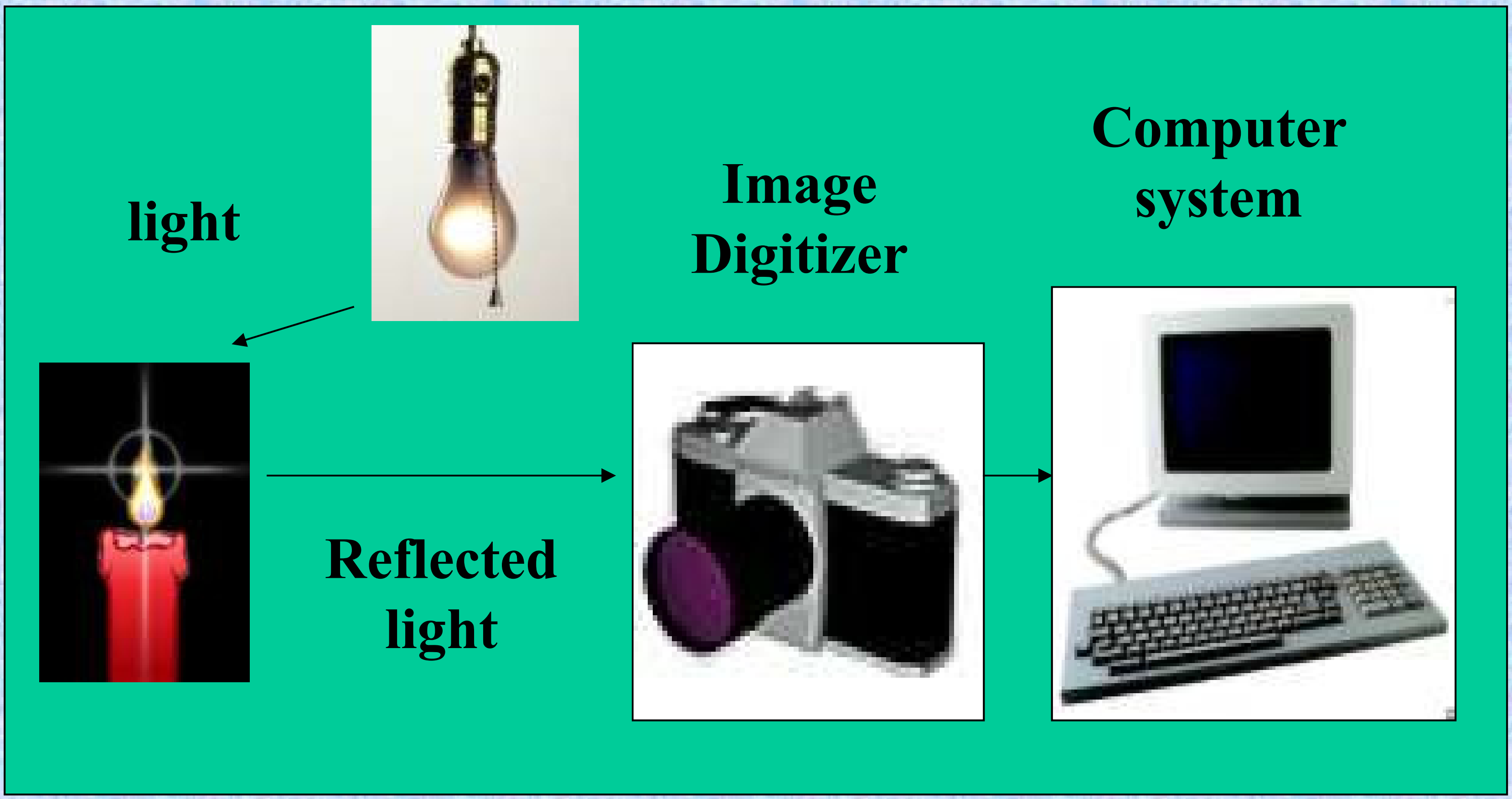
+/- 05 marks variation at any part;
To be finalized well before end sem exam.

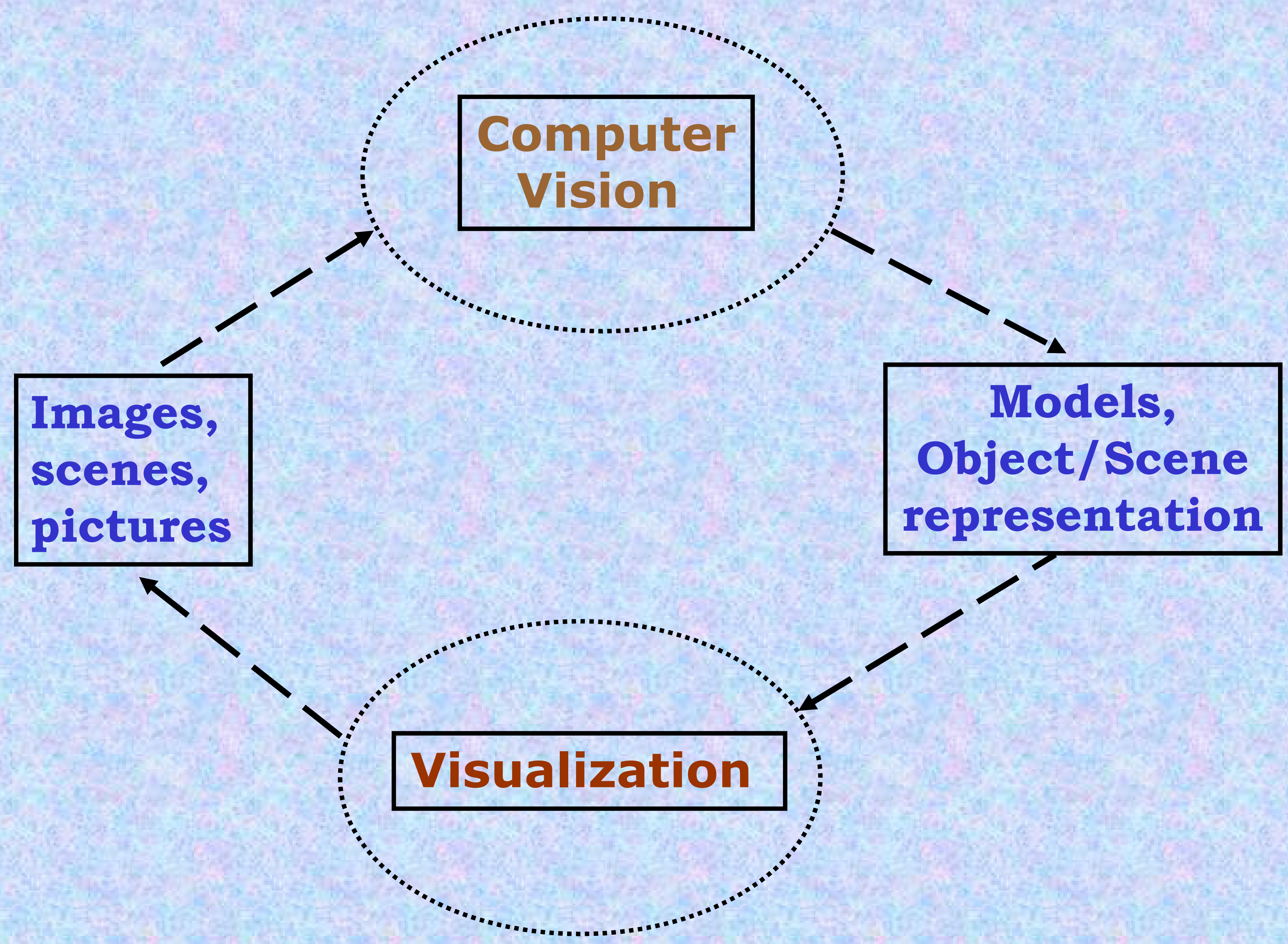
Human Vision System (HVS) Vs. Computer Vision System (CVS)



The Optics of the eye

A computer Vision System (CVS)



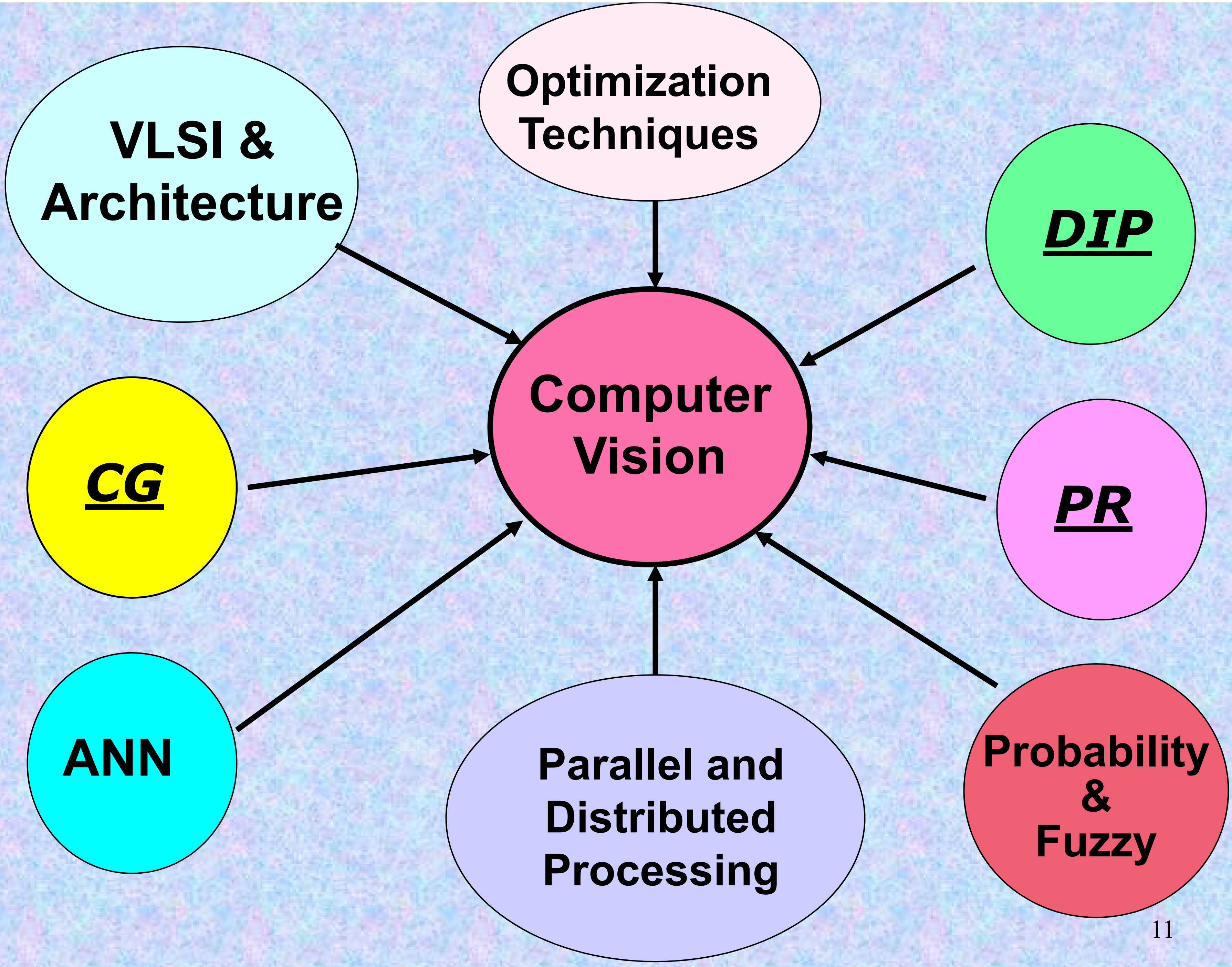

Computer Vision
Models, Object/Scene representation
Visualization
Images, scenes, pictures

Computer Vision is an area of work, which is a combination of concepts, techniques and ideas from Digital Image Processing, Pattern Recognition, Artificial Intelligence and Computer Graphics.

Majority of the tasks in the fields of Digital Image Processing or Computer Vision deals with the process of understanding or deriving the scene information or description, from the input scene (digital image/s). The methods used to solve a problem in digital image processing depends on the application domain and nature of data being analyzed.

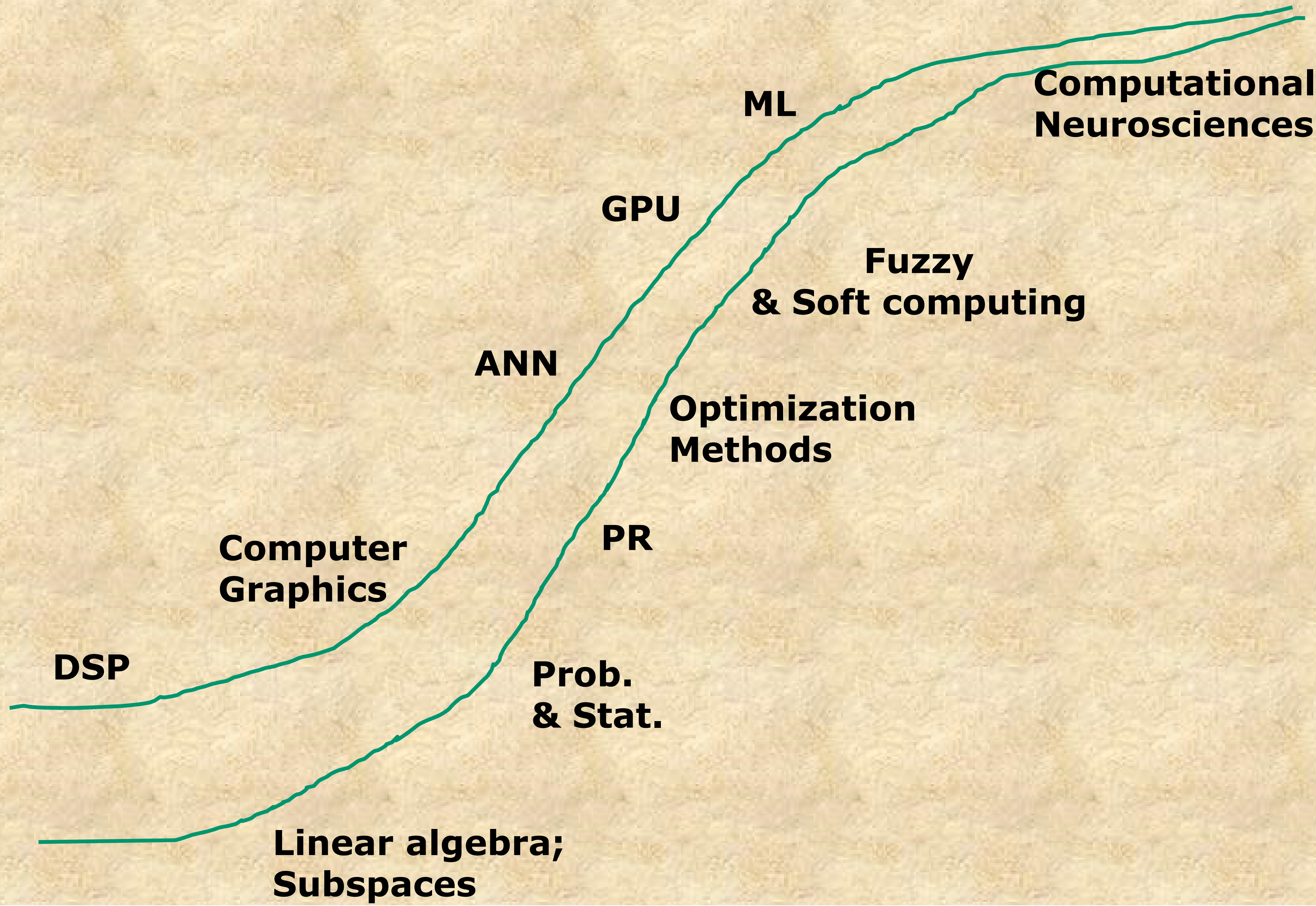
Analysis of two-dimensional pictures are generally not applicable of processing three-dimensional scenes, and vice-versa. The choice of processing, techniques and methods and 'features' to be used for a particular application is made after some amount of trial and error, and hence experience in handling images is crucial in most of these cases.

For example, analysis of remote sensed or satellite imagery involves techniques based on classification or analysis of texture imagery. These techniques are not useful for analyzing optical images of indoor or outdoor scenes.



Optimization Techniques
VLSI & Architecture
DIP
Computer Vision
CG
PR
ANN
Parallel and Distributed Processing
Probability & Fuzzy

The Developmental Pathway of Computational Vision Technology



Digital Image processing is in many cases concerned with taking one array of pixels as input and producing another array of pixels as output which in some way represents an improvement to the original array.

Purpose:

1. **Improvement of Pictorial Information**
   - improve the contrast of the image,
   - remove noise,
   - remove blurring caused by movement of the camera during image acquisition,
   - it may correct for geometrical distortions caused by the lens.

2. **Automatic Machine perception** (termed Computer Vision, Pattern Recognition or Visual Perception) for intelligent interpretation of scenes or pictures.

Elements of a Digital Image Processing System

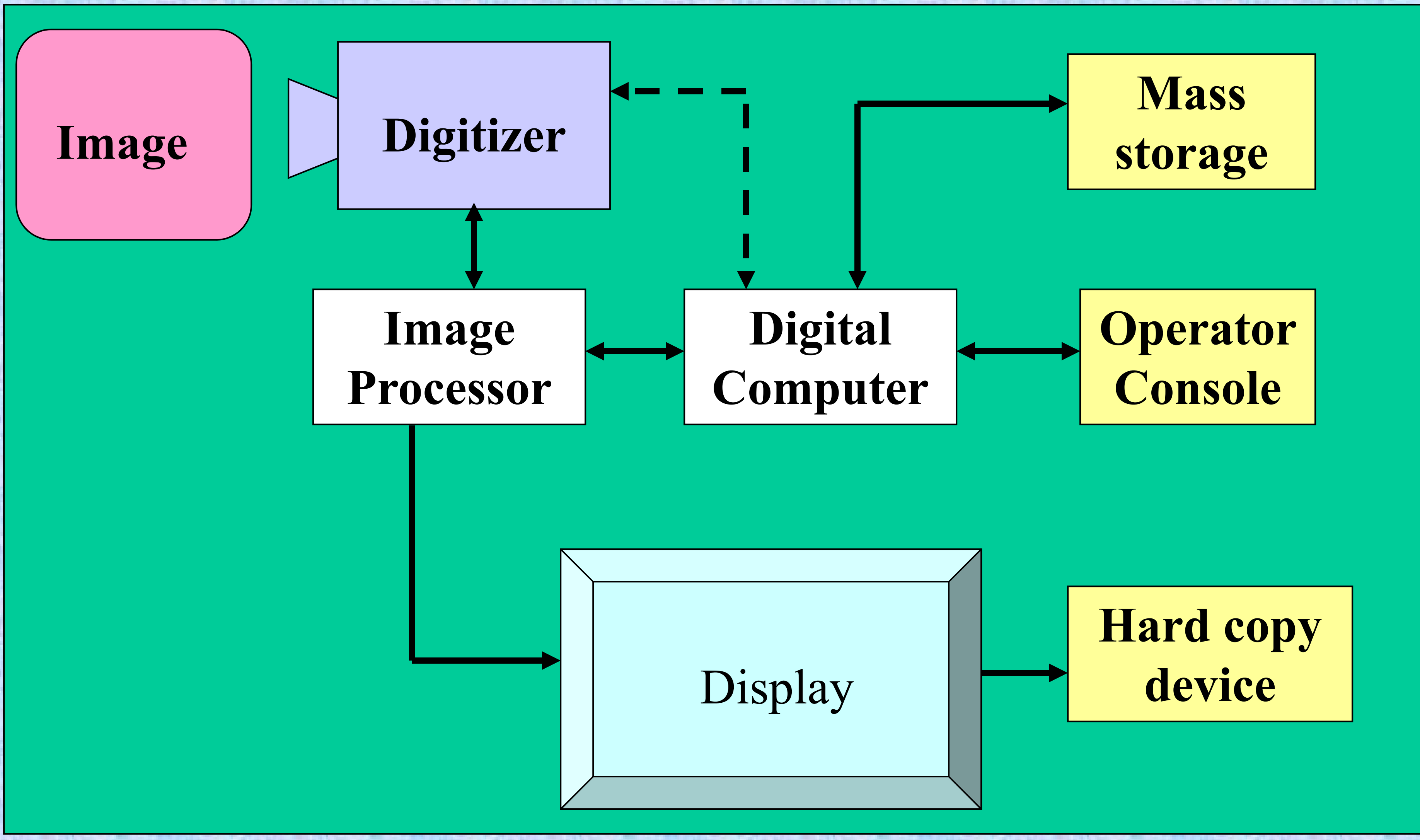
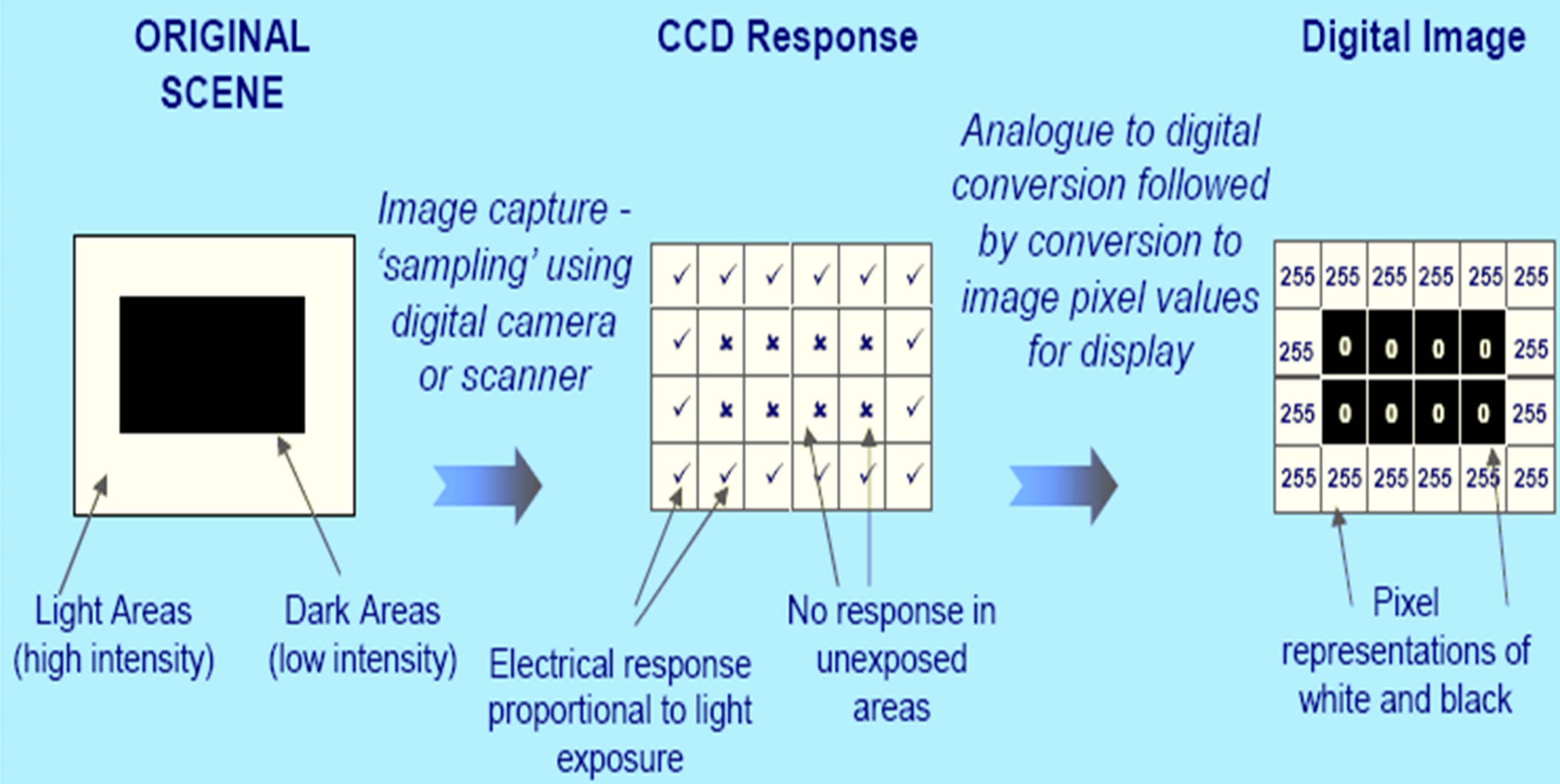


Image processors: Consists of set of hardware modules that perform 4 basic functions:

- **Image acquisition: frame grabber**
- **Storage: frame buffer**
- **Low-level processing: specialized hardware device designed to perform Arithmetic Logic operations on pixels in parallel**
- **Display: read from image memory (frame buffer) and convert to analog video signal**

- **Digitizers:** Converts image into numerical representation suitable for input to a digital computer
- **Digital Computers:** Interfaced with the image processor to provide versatility and ease of programming.
- **Storage Devices:** For bulk storage. e.g:- Magnetic disks, magnetic tapes, optical disks
- **Display and Recording devices :** Monochrome and Color Television monitors, CRT, Laser printers, heat-sensitive paper devices, and ink spray systems.

Image acquisition using a CCD camera



A digital Image

Image is an array of integers: f(x,y) ε {0,1,….,I_{max}-1},
where, x,y ε {0,1,…..,N-1}

- N is the resolution of the image and I_{max} is the level of discretized brightness value
- Larger the value of N, more is the clarity of the picture (larger resolution), but more data to be analyzed in the image
- If the image is a gray-level (8-bit per pixel - termed raw, gray) image, then it requires N2 Bytes for storage
- If the image is color - RGB, each pixel requires 3 Bytes of storage space.

Image Size (resolution)	Storage space required	
	Raw - Gray	Color (RGB)
64*64	4K	12K
256*256	64K	192K
512*512	256K	768K

A ***digital image*** is a two-dimensional (3-D image is called range data) array of intensity values, f(x, y), which represents 2-D intensity function discretized both in spatial coordinates (**spatial sampling**) and brightness (**quantization**) values.

The elements of such an array are called **pixels** (picture elements).

The storage requirement for an image depends on the **spatial resolution** and number of bits necessary for **pixel quantization**.

The processing of an image depends on the application domain and the methodology used to solve a problem. There exists four broad categories of tasks in digital image processing:

(i) Compression, (ii) Segmentation,
(iii) Recognition and (iv) motion.

Segmentation deals with the process of fragmenting the image into homogeneous meaningful parts, regions or sub-images. Segmentation is generally based on the analysis of the histogram of images using gray level values as features. Other features used are edges or lines, colors and textures.

Recognition deals with identification or classification of objects in an image for the purpose of interpretation or identification. Recognition is based on models, which represent an object. A system is trained (using HMM, GMM, ANN etc.) to learn or store the models, based on training samples. The test data is then matched with all such models to identify the object with a certain measure of confidence.

Compression involves methodologies for efficient storage and retrieval of image data, which occupies large disk space. Typical methods are, JPEG-based, Wavelet based, Huffman Coding, Run length coding etc. for still images and MPEG-I, II, IV & VII for digital video or sequence of frames.

Motion analysis (or dynamic scene analysis) involves techniques for the purpose of tracking and estimation of the path of movement of object/s from a sequence of frames (digital video). Methods for dynamic scene analysis are based on (i) tracking, (ii) obtaining correspondence between frames and then (iii) estimating the motion parameters and (iv) structure of moving objects. Typical methods for analysis are based on optical flow, iterative Kalman filter and Newton/Euler's equations of dynamics.

There are generally three main categories of tasks involved in a complete computer vision system. They are:

- ***Low level processing***: Involves image processing tasks in which the quality of the image is improved for the benefit of human observers and higher level routines to perform better.

- ***Intermediate level processing***: Involves the processes of feature extraction and pattern detection tasks. The algorithms used here are chosen and tuned in a manner as may be required to assist the final tasks of high level vision.

- ***High level vision***: Involves autonomous interpretation of scenes for pattern classification, recognition and identification of objects in the scenes as well as any other information required for human understanding.

A top down approach, rather than a bottom-up approach is used in the design of these systems in many applications. The methods used to solve a problem in digital image processing depends on the application domain and nature of data being analyzed.

Different fields of applications include:

- Character Recognition,
- Document processing,
- Commercial (signature & seal verification) application,
- ***Biometry and Forensic*** ***(authentication: recognition and verification of persons using face, palm & fingerprint),***
- **Pose** and gesture identification,
- Automatic inspection of industrial products,
- Industrial process monitoring,
- Biomedical Engg. (Diagnosis and surgery),
- ***Military surveillance and target identification,***
- Navigation and mobility (for robots and unmanned vehicles - land, air and underwater),
- ***Remote sensing (using satellite imagery),***
- ***GIS***
- Safety and security (night vision),
- Traffic monitoring,
- *Sports (training and incident analysis)*
- VLDB (organization and retrieval)
- Entertainment and virtual reality.

TARGETED INDUSTRIAL APPLICATIONS

Intelligent Traffic Control

Anti-forging Stamps

Card Counting Systems

Drive Quality Test

Camera Flame Detection

CCTV Fog Penetration

Key Image Search/Index

Security Monitoring

Robust Shadow Detection

Vehicle Segmentation

Visual Tracking Systems

Illegal content (adult) Filter

Scratch Detection

Smart Traffic Monitoring

Vehicle Categorization

Vehicle Wheel alignment

Number Plate Identification

Referrals for Line calls

Different categories of work being done in CV, to solve problems:

2-D image analysis – segmentation, target detection, matching, CBIR;

3-D multi-camera calibration; Correspondence and stereo; Reconstruction of 3-D Objects and surfaces;

Pattern Recognition for Objects, scenes;

Video and motion analysis; Video analytics; CBVR; Compression;

Feature extraction: Canny, GHT, Snakes, DWT, Corners, SIFT, GLOH, LESH;

Multi-sensor data, Decision and feature fusion;

Image and Video-based Rendering;

Steganography and Watermarking;

The various sub-categories of technology in these related fields are:

- *image enhancement,*
- *image restoration and filtering,*
- *representation and description,*
- *feature extraction,*
- *image segmentation,*
- *image matching,*
- color image processing,
- image synthesis,
- image representation,
- *image reconstruction*
- *range data processing,*
- *stereo image processing*
- *computational geometry,*
- *image morphology,*
- *artificial neural networks,*
- *Neuro-fuzzy techniques,*
- *computational geometry,*
- *parallel architectures & algorithms.*

What is CVPR ?

http://cvpr2016.thecvf.com/;

http://www.pamitc.org/cvpr16/index.php;

https://scholar.google.com/citations?view_op=top_venues&hl=en

	Publication (2016)	h5-index	h5-median
1.	Nature	377	529
2.	The New England Journal of Medicine	328	520
3.	Science	316	446
4.	The Lancet	258	415
5.	Cell	216	330
25.	PLoS ONE	161	210
41.	Physical Review D	138	178
51.	Nature Immunology	129	218
52.	Nature Reviews Molecular Cell Biology	129	216
53.	The American Economic Review	129	196
54.	Hepatology	129	181
55.	IEEE Conference on Computer Vision and Pattern Recognition, CVPR	128	203
57.	Immunity	128	177
58.	Nature Reviews Immunology	127	214
60.	Nature Methods	125	185
61.	Nature Neuroscience	125	175
100.	IEEE Transactions on Industrial Electronics	109	147

CVPR-16 - CFP

- 3D Computer Vision
- Action Recognition
- Biometrics
- Big Data, Large Scale Methods
- Computational Photography,
- Sensing, and Display
- Convolutional Neural Networks and Deep Learning
- Document Analysis
- Face and Gesture
- Kinect/3D
- Low-level Vision, Image Processing
- Medical, Biological and Cell
- Microscopy Image Analysis
- Motion and Tracking
- Optimization Methods
- Performance Evaluation and Data Sets
- Physics-based Vision and Shape From X
- Recognition: Detection, Categorization, Indexing, Matching
- Segmentation, Grouping, and Shape Representation
- Statistical Methods and Learning
- Video: Events, Activities, and Surveillance
- Vision for Graphics
- Vision for Robotics
- Vision for Web

Few DEMOS and ILLUSTRATIONS

Courtesy: students of VPLAB - CSE-IITM

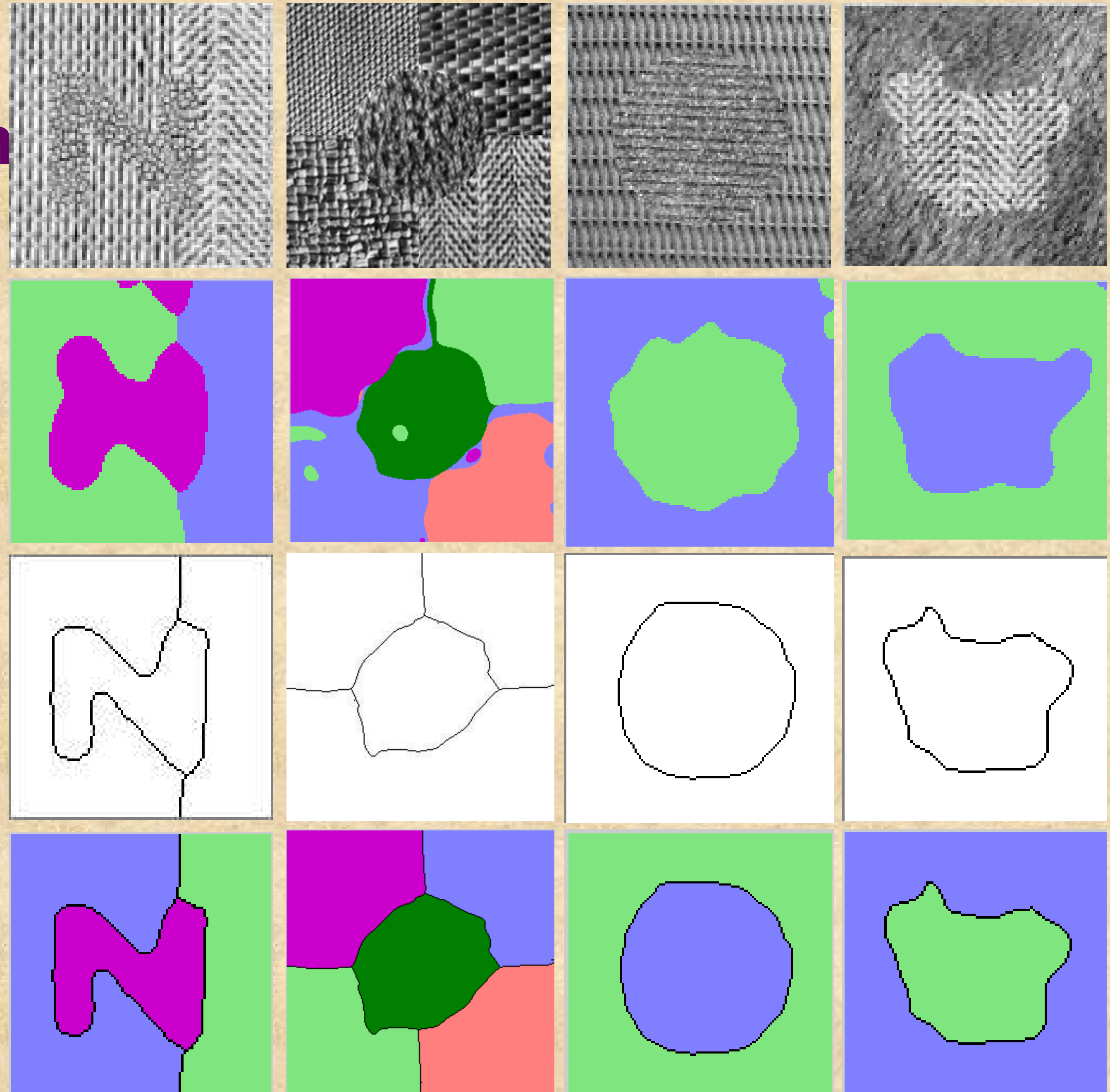
Results of Segmentation

Input Image

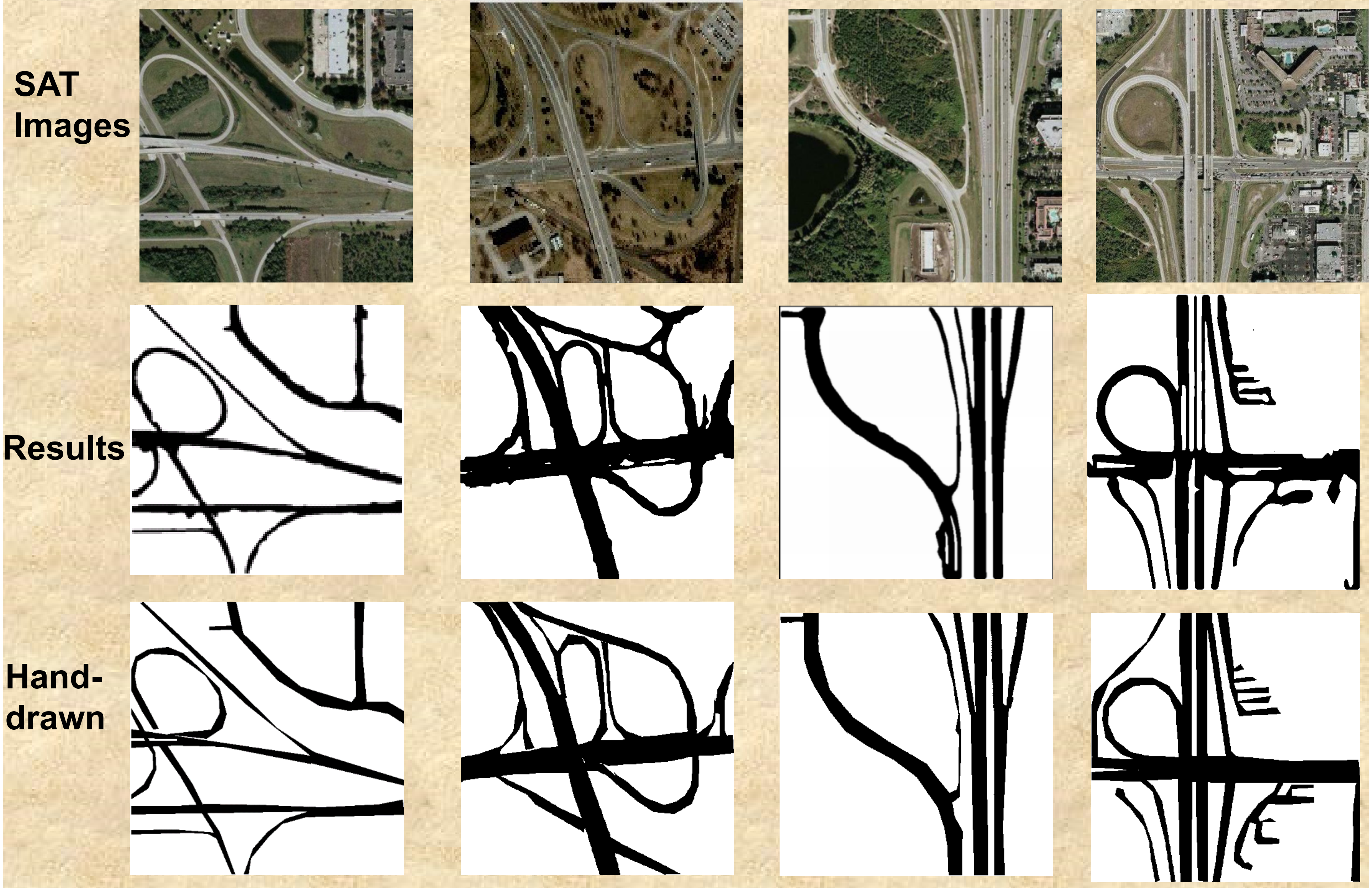
Segmented map before integration

Edge map before integration

Segmented map and Edge map after integration



Road extraction from Satellite Images



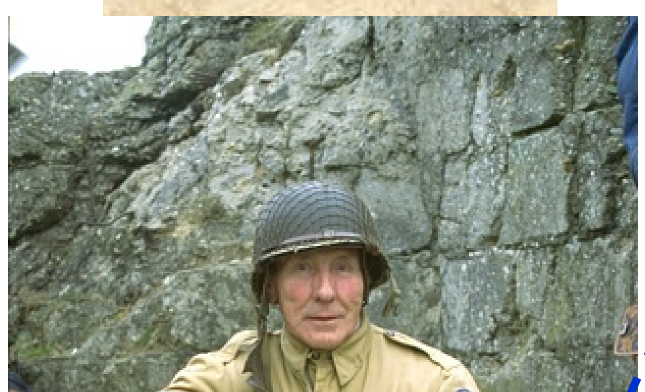
SNAKECUT – Extraction of a Foreground Object with holes

Our proposed approach for segmentation of an object with a hole, using a combination of (i) Active Contour and (ii) GrabCut

- **Here, objective is to crop the soldier from the input image**
- **Cropped image should not contain any part of the background**

Snake Output

GrabCut Output

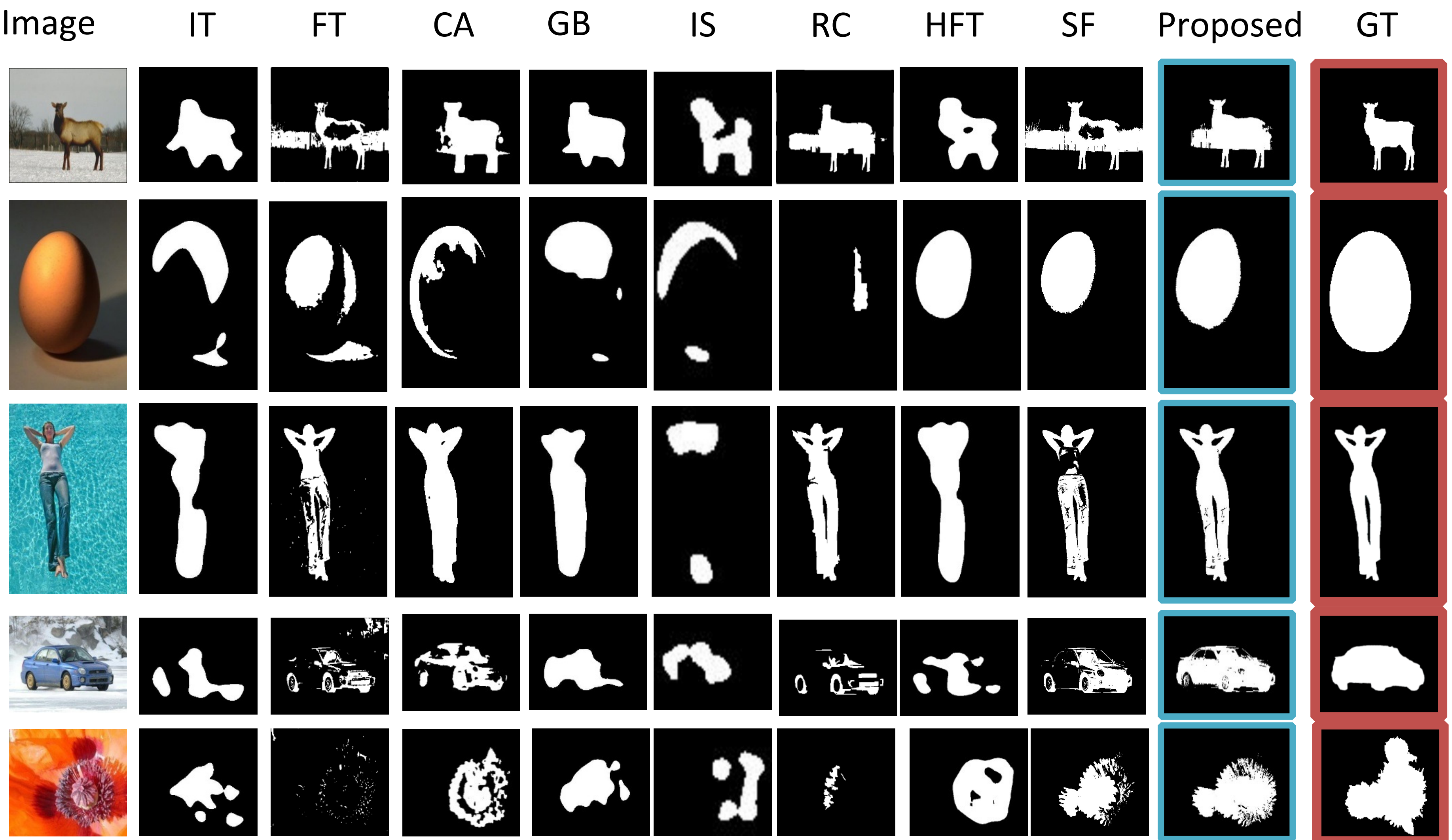


SnakeCut Output

Object Extraction From an Image

Method 1

Unsupervised Saliency

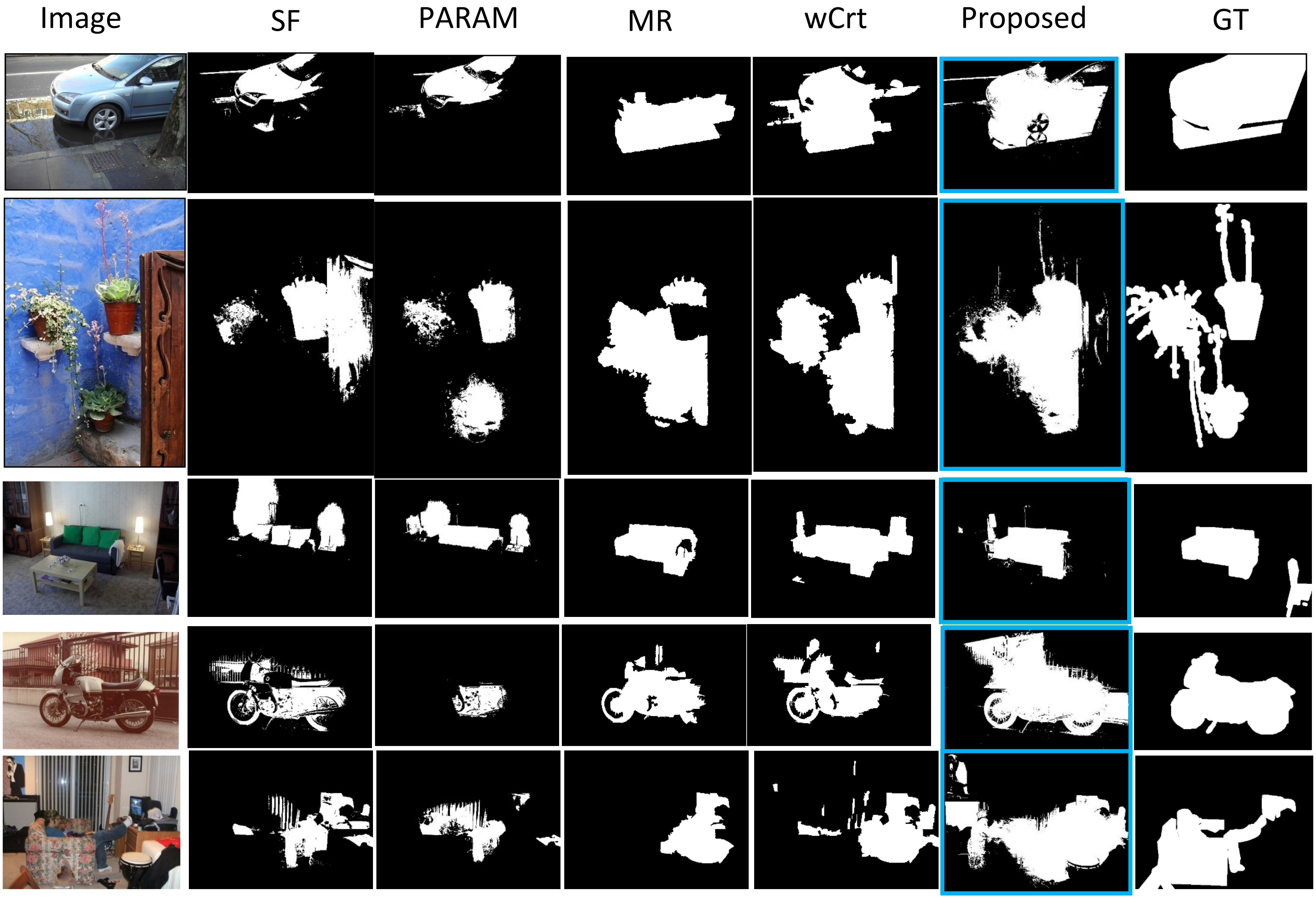


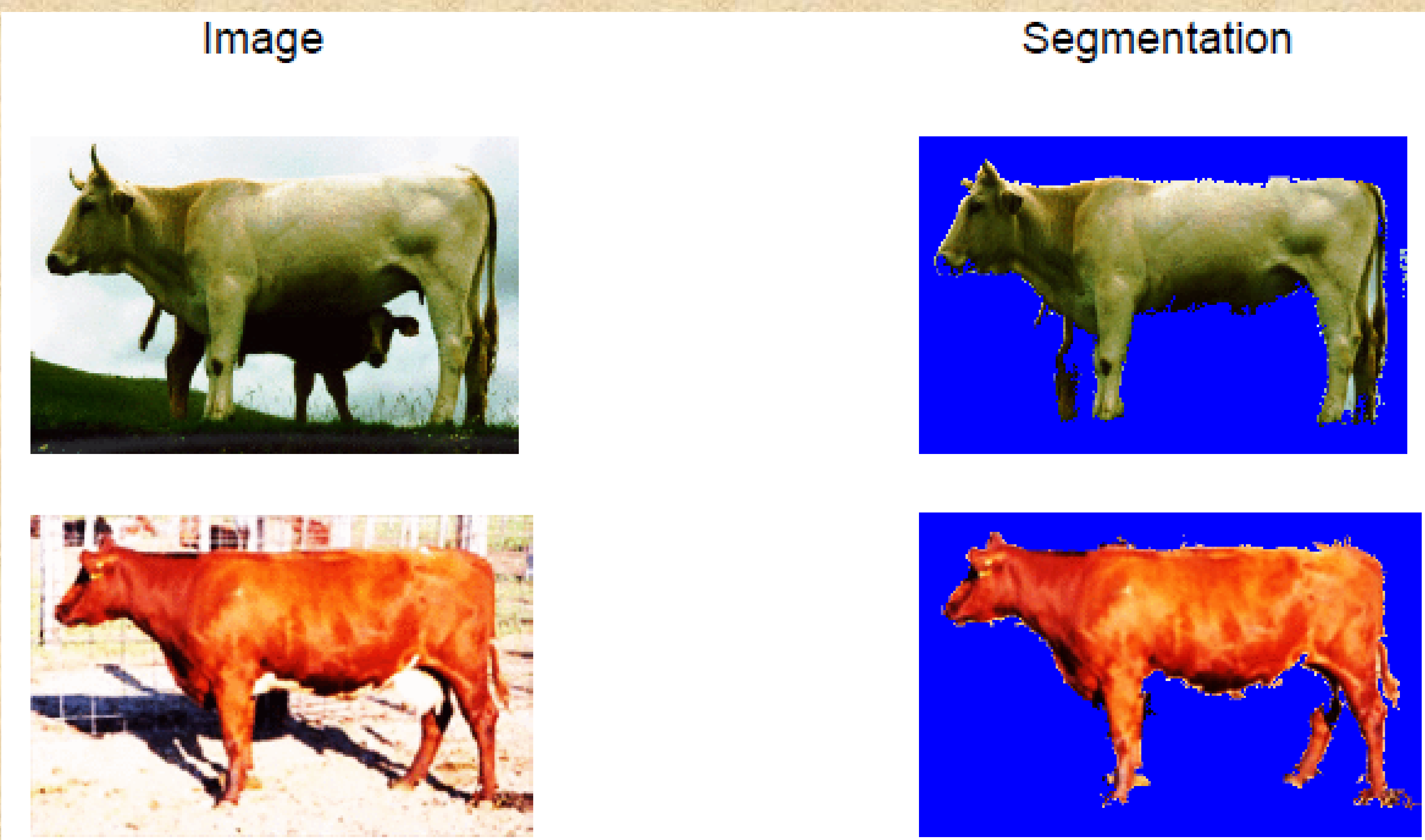
Images from MSRA B 5000 image Dataset

http://research.microsoft.com/en-us/um/people/jiansun/SalientObject/salient object.htm

Method 2

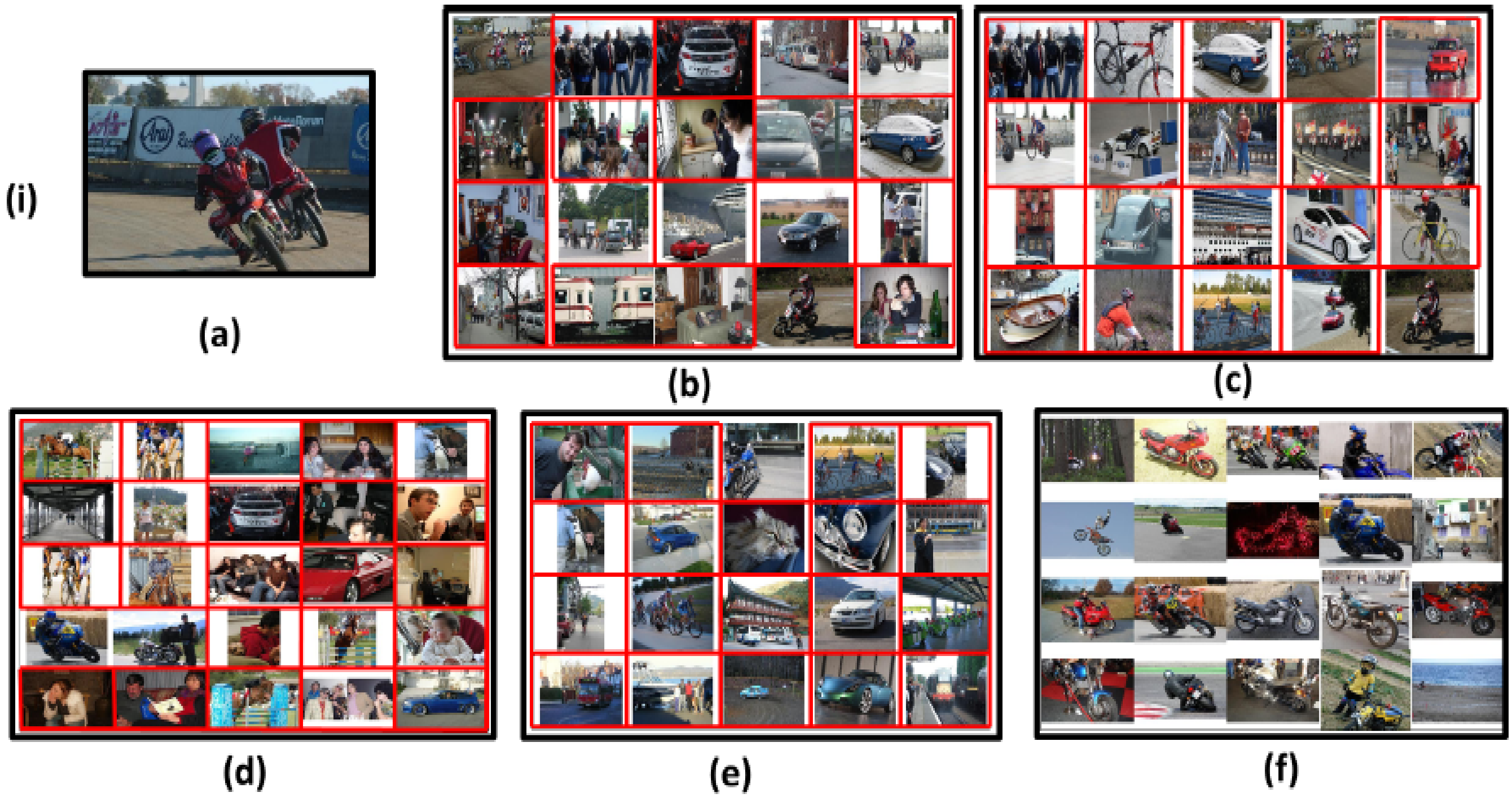
Visual Results on PASCAL





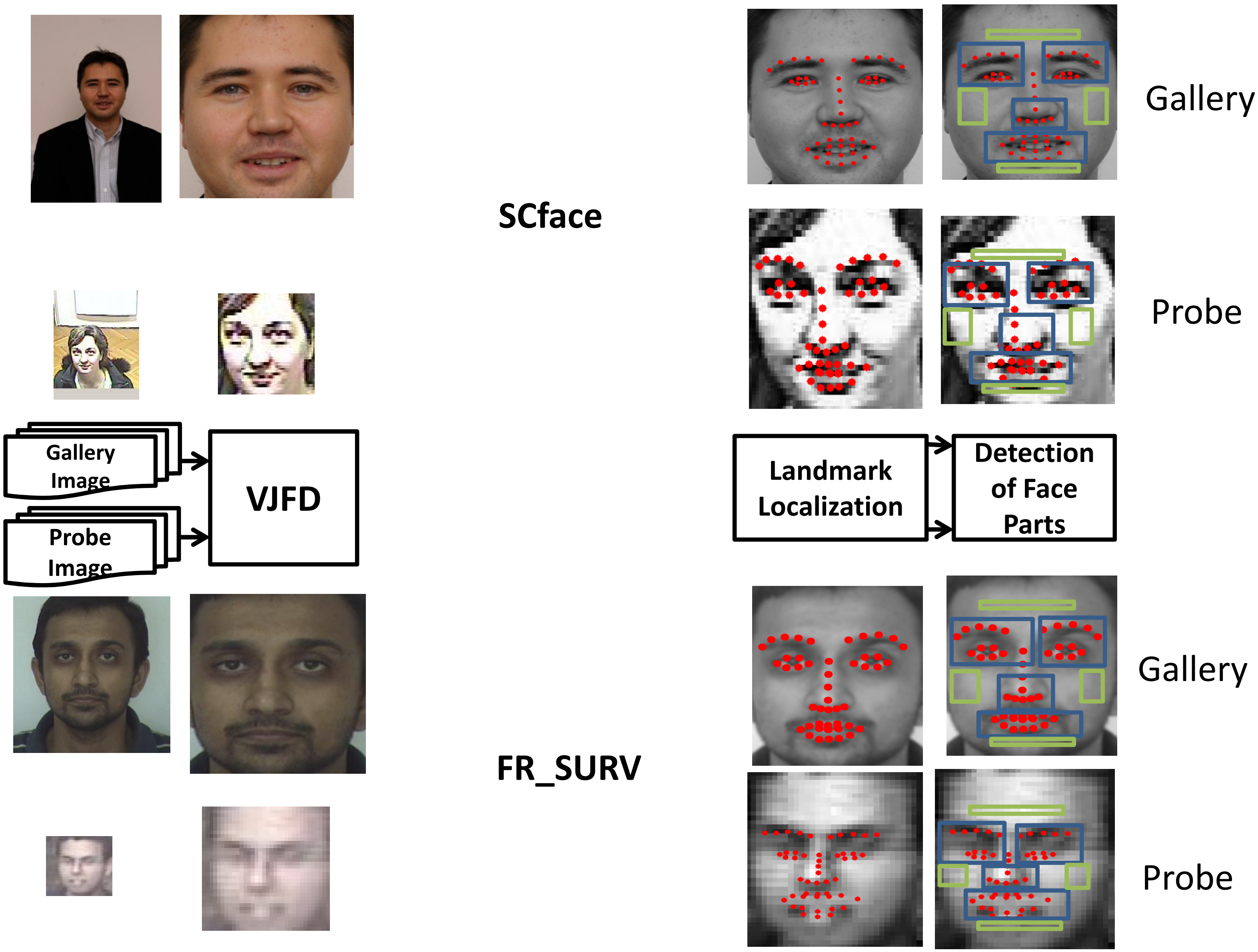
Object Detection or segmentation –
involves object Detection and recognition modules

Smart CBIR - Retrieval Results



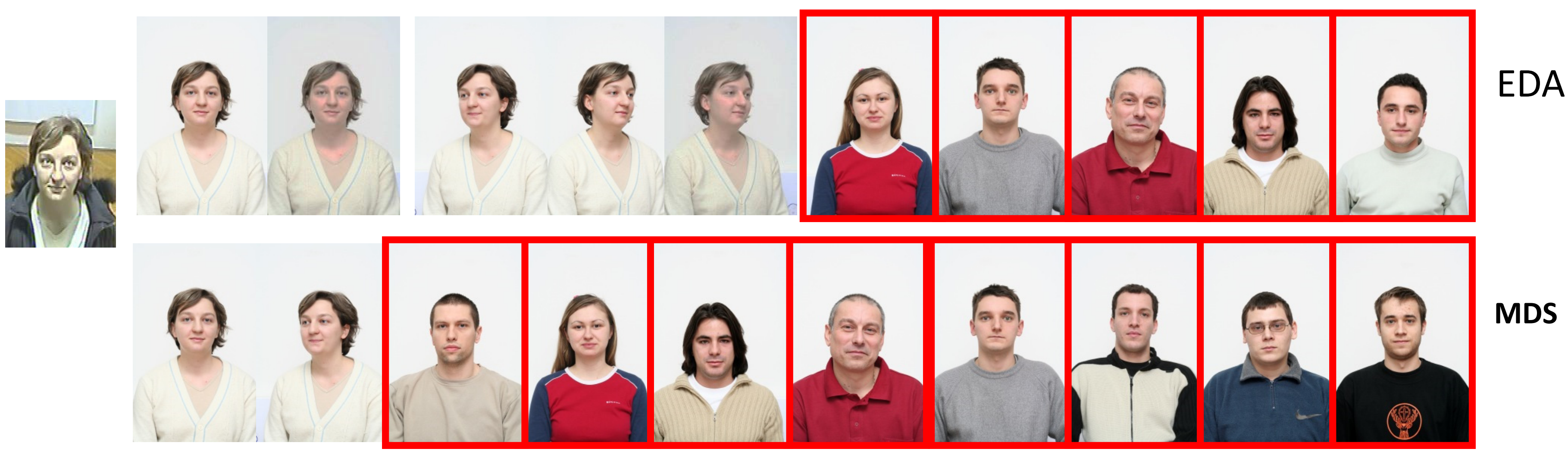
Results of top 20 image retrievals (arranged in row-major order) shown for visual comparative study, using: (a) query image from the PASCAL datasets; (b) MTH (2010); (c) MSD (2011), (d) SLAR (2012); (e) CDH (2013); and (f) our proposed RADAR framework. Erroneous results are highlighted using a red template

Intermediate stages of Face Processing



Top 10 Retrievals for SCface

Rank – ordered Retrieval Results →



Erroneous retrievals are marked by red templates.

Top 10 Retrievals for FR_SURV

Query

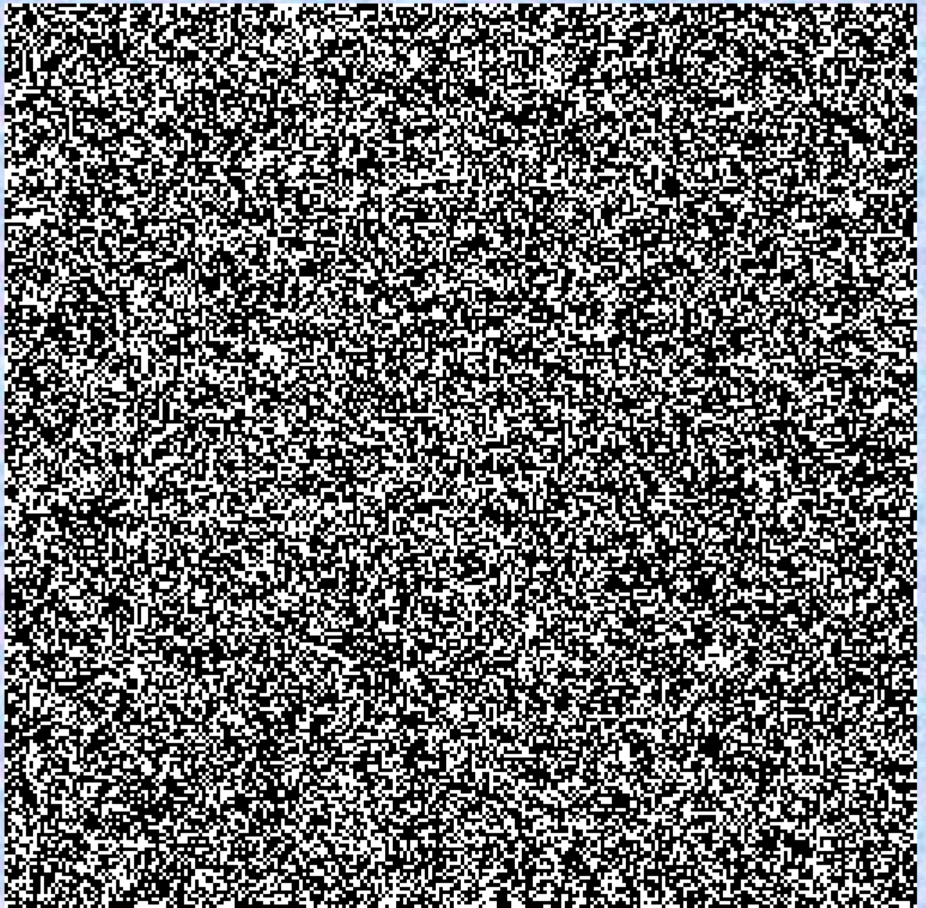
Rank – ordered Retrieval Results →

Erroneous retrievals are marked by red templates.

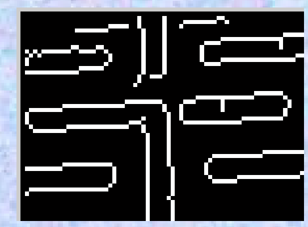
SIFT : Result

Object detection

The Problem Definition



IMRN



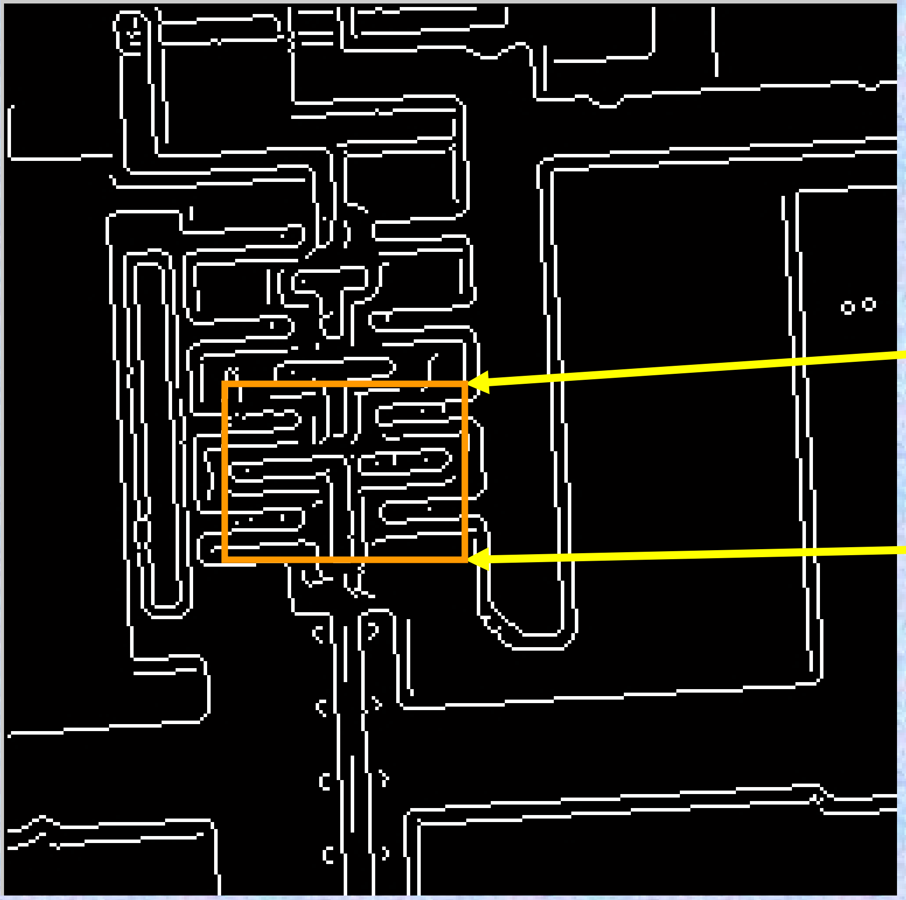
IMT

Given a bitmap template (IMT) and a noisy bitmap image IMRN which contains IMT (believe me):
FIND OUT the location of IMT in IMRN !

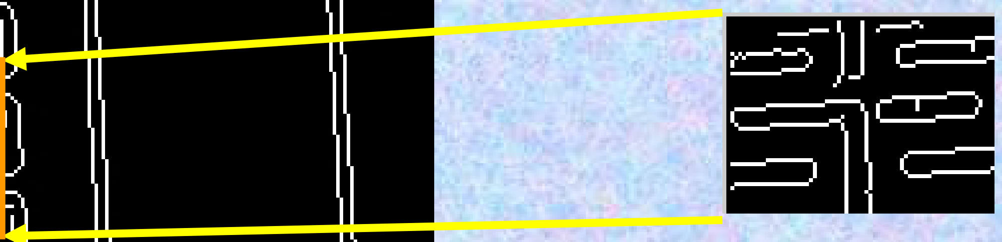
Go to the next page for more:

Problem explanation for pessimists.

IMR



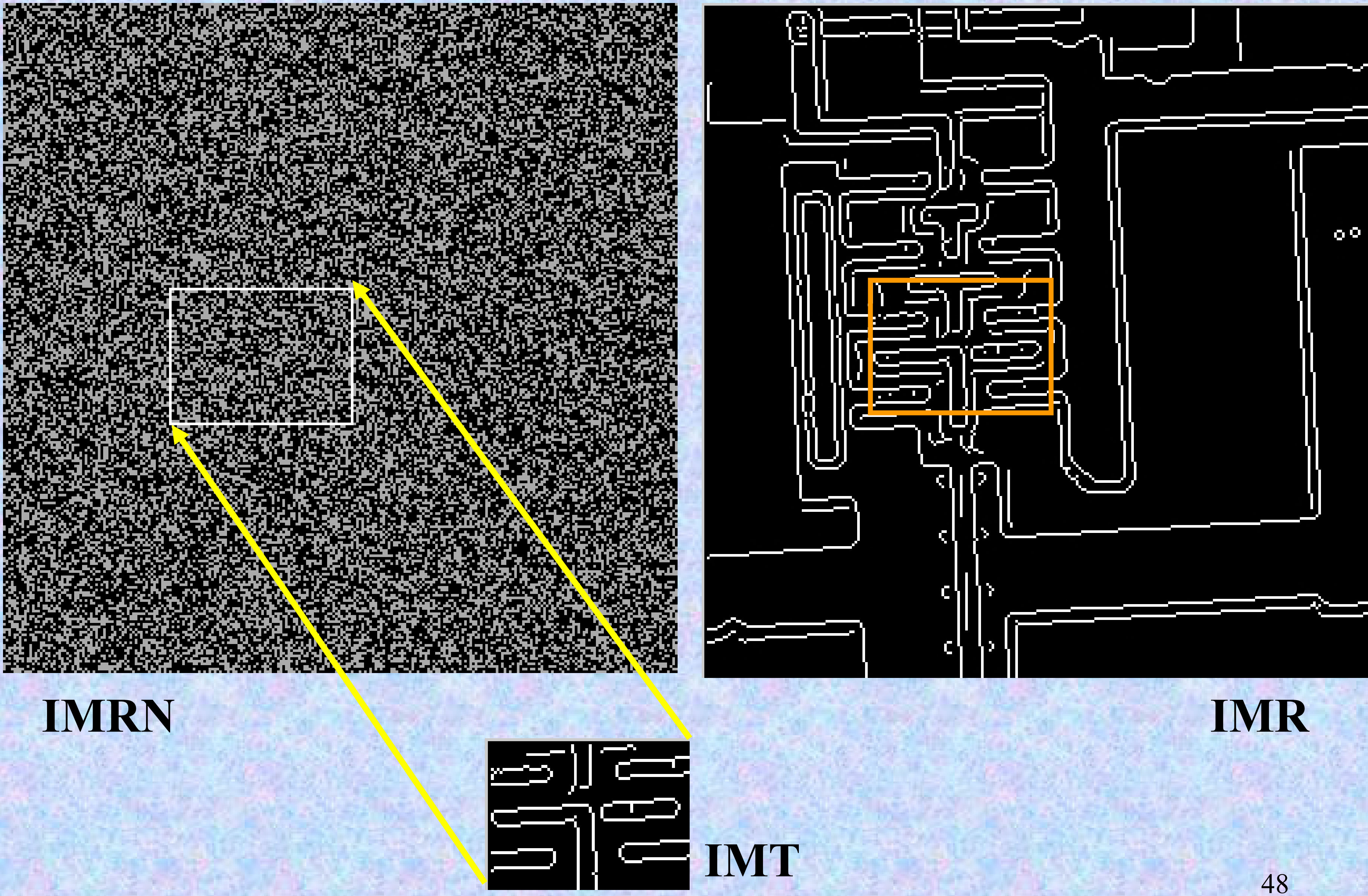
IMT



- IMRN (in previous page) is obtained by adding a large level of "Salt and Pepper" noise onto IMR bitmap image.

- IMT is also obtained from IMR as shown above.

The RESULT beats the human EYE



IEEE TRANSACTIONS ON COMMUNICATIONS,
VOL. COM-34, NO. 11, NOVEMBER 1986

Classified Vector Quantization of Images

BHASKAR RAMAMURTHI
& Allen G.

Thank you